Developing Scientific Research Proposals (Grant Writing)

2003 Epidemiology and Biostatistics Summer Session

Alan Kristal, Dr.P.H.

Session 7

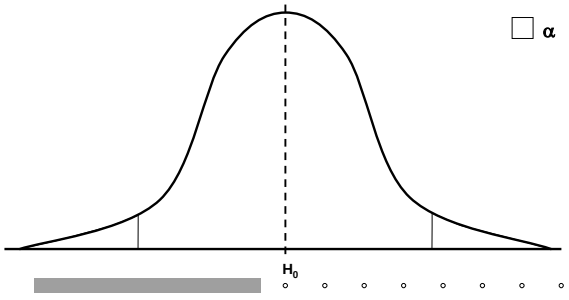
Statistical Power

Power

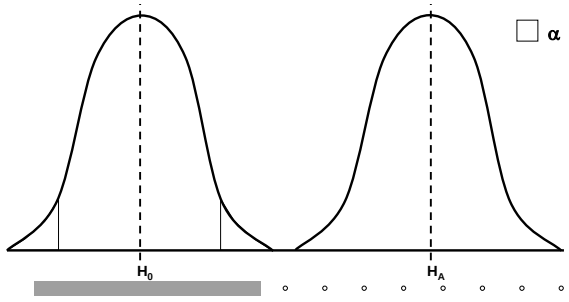
Purpose

- Demonstrate that at the completion of data collection there will be sufficient numbers of observations to test primary specific aims.
- Recommended reference: Kelsey JL, Whittmore AS, Evans AS, Thompson WD. Methods in Observational Epidemiology. Oxford University Press, 2nd ed. 1996

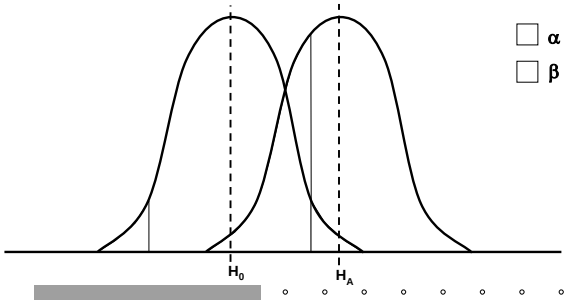
Sampling Distribution of Null Value of Measure of Association



Sampling Distribution of Extreme Non-Null Value of Measure of Association



Sampling Distribution of Non-Null Value of Measure of Association



Components of Statistical Power

Power is a function of four parameters

- Effect size
- Sample size
- Probability of a non-significant test when H_o should be rejected (β error)
- Probability of a significant test when H_o should not be rejected (α error)

Components of Statistical Power

Effect Size

- Magnitude of intervention effect
- Relative risk
- Odds ratio
- Hazard ratio
- Shared variance (correlation or regression)

Components of Statistical Power

Sample size

- Number of cases and controls (case/control study)
- Number of participants (cohort, experiment)

Components of Statistical Power

Probability of significant test when H_o should not be rejected

- Alpha error (α)
- One sided vs. two sided

Components of Statistical Power

Probability of non-significant test when H_o should be rejected

- Beta error (β)
- Power = 1- β

Describing Power

Alpha error is almost always set at 0.05 (two sided). Thus, power sections usually take one of the following three approaches:

1. Solve for minimally detectable effect size
2. Solve for number of study participants
3. Solve for power

Describing Power

Solve for minimally detectable effect size

- You have a limited number of participants available for study. For example, cases of newly diagnosed breast cancer in King County.
- You set power at 80%.
- What relative risk could you detect.

Power Table - Fixed Sample Size

Minimally Detectable Effect in Nested Case-Cohort Analysis

Specific Aims	Title	Cases (n)	Cohort or Controls (n)	Minimal Detectable RR
Specific Aim1 (a), (b) and 2 (a) and (b)	Diet and Prostate Cancer	700	1400	0.70
Specific Aim 1 (c)	Serum Fatty Acids and Prostate Cancer	700	1400	0.40
Specific Aim A.2 (c)	Serum Micronutrients and Prostate Cancer	700	1400	0.60

Describing Power

Solve for number of participants

- You know or hypothesize an effect size. For example, you determine that a 25% reduction in risk is worth detecting.
- You set power at 90%.
- How many participants do you need?

Power Table - Intervention Trial

Trial of Behavioral Intervention for Dietary Change

Instrument	Measure	Smallest Meaningful Intervention Effect Between Arms	Power (1-B)	Minimum Detectable Difference with 90% Power
Principal Endpoints				
24-Hour Diet Recall	Fat (% En)	2 percentage points	.90	2.0
	Fiber (g/ 1000 Kcal)	2 g	>.95	1.8

Power Table - Intervention Trial

Trial of Behavioral Intervention for Dietary Change

Instrument	Measure	Smallest Meaningful Intervention Effect Between Arms	Power (1-B)	Minimum Detectable Difference 90% Power
Secondary Endpoints				
Fat and Fiber Behavior (FFB)	Fat Scale Fiber Scale	0.14 units 0.19 units	> .95 > .95	.057 .064
Stage of Change	Percent moving into action stage	Fat: 18 percentage points Fiber: 17 percentage points	.94 > .95	.14 .11

Describing Power

Solve for power

- You have a limited number of participants available.
- You declare a minimally detectable effect size.
- How much power would have to detect effect sizes at least this large?

Power Table - Power to Detect Effects

Case-Control Study of Medication Use and Prostate Cancer Risk

Odds ratio	Prevalence of exposure among controls					
	0.05	0.10	0.15	0.20	0.30	0.40
0.5	0.79	0.98	> 99	> 99	> 99	> 99
0.6	0.56	0.85	0.95	> 99	> 99	> 99
0.7	0.32	0.57	0.74	0.84	0.93	0.96
0.8	0.15	0.27	0.38	0.46	0.59	0.66
1.3	0.24	0.43	0.57	0.66	0.77	0.82
1.4	0.39	0.65	0.80	0.87	0.94	0.96
1.5	0.54	0.82	0.93	0.97	0.99	0.99
1.8	0.89	0.99	> 99	> 99	> 99	> 99
2.0	0.97	> 99	> 99	> 99	> 99	> 99

* Based on a total sample size of 1000 cases, 1000 controls; alpha=0.05, 2-tailed.

Estimating Effect Size

- Decreased incidence in treatment vs. placebo study arm
- Difference in incidence in persons exposed vs. not exposed
- Relative risks/odds ratios comparing persons in highest to lowest exposure category

Estimating Effect Size

How to determine effect size *(in order or preference)*

- Pilot study
- Clinically meaningful
- Significant at population level
- By analogy to similar exposure to similar disease
- Ideology (standards in one's field)

Assumptions for Power Calculations

Cohort Studies

- Incidence rate
- Proportion of population exposed
- Variability of exposure in population

Assumptions for Power Calculations

Case-Control Studies

- Number of Cases and Controls
- Matched vs. unmatched analyses
- Proportion of exposed controls

Assumptions for Power Calculations

Intervention Trials

- Adherence
- Drop-out
- Drop-in
- Incidence rate in contrast (untreated) group

How Much Power is Enough?

- For observation studies, try to achieve at least 80% power (1-β), with two-sided alpha error at 5% ($p<0.05$).
- For experiments with disease outcomes, try to achieve 90% power with two-sided alpha error at 5%.

How Much Power is Enough?

- For experiments with non-disease outcomes, balance costs and societal importance. Is the cost of an effect not detected due to low power more the cost of increasing sample size?

Alpha-Error and Multiple Testing

- Do not ignore!
- Adjust α if feasible (especially for intervention trial with multiple endpoints)
- Site a-priori hypotheses as those not needing protection from multiple testing
- Move as many tests to secondary aims as possible (power less critical for secondary aims)

Methods - Common Criticisms

Design

- Lack of adequate control group
- Inadequate rationale for selection of control or comparison group
- Model not appropriate for hypothesis

Methods - Common Criticisms

Participants

- Population not representative
- Insufficient evidence for recruiting sufficient numbers
- Overly optimistic estimates of participation, adherence, or drop out

Methods - Common Criticisms

Assessments

- Measures not validated
- Over-reliance on self-report
- Unrealistic participant burden
- Fishing expedition
- Data collection protocol not comparable across time or study groups

Methods - Common Criticisms

Intervention

- No conceptual framework
- No pilot data on efficacy
- Too complex
- Unrealistic participant burden
- Differential attrition across study arms
- Dangerous

Methods - Common Criticisms

Treatment Trials

- Treatment not blinded
- Inadequate or no assessment of compliance
- Dose/protocol not justified
- No control for contamination/drop-ins

Methods-Common Criticisms

Data analysis

- Power estimates missing for primary specific aims
- Analysis plan does not match study design
- Analysis proposed but not described
- Analysis plan is vague - not clear how hypotheses will tested